

El principio de representatividad y las nuevas técnicas de muestreo

Andrés Gutiérrez

Universidad Santo Tomás
Agencia Presidencial para la Acción Social

2011

1 Motivación

La fábula de Basú

La historia incompleta

2 Introducción

Definiciones básicas

Un poco de historia

3 Diseño de muestreo

4 Estimación

La regla de oro

5 Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Discusión

6 Comparación empírica

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

El dueño de un circo está planeando transportar sus 50 elefantes adultos, para este propósito él necesita una buena estimación del peso total de los elefantes. Como pesar un elefante es una tarea muy incómoda, el dueño del circo quiere estimar el peso total pesando sólo un elefante. ¿Cuál elefante debería pesar? El dueño del circo decide echar un vistazo a sus registros y descubre una antigua lista de los pesos de los elefantes, elaborada hace tres años.

Él encuentra que tres años atrás Sambo, un elefante mediano, era el promedio (en peso) de su manada. El dueño del circo verifica la información con el entrenador quien le asegura que Sambo todavía puede considerarse como el promedio de la manada.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Así, el dueño del circo planea pesar a Sambo y tomar a 50 y (donde y es el peso de Sambo) como una estimación del peso total ($t_y = y_1 + \dots + y_{50}$) de la manada. Pero el estadístico del circo se aterra al conocer el plan de muestreo del dueño (con un diseño de muestreo no probabilístico).

- ¿Cómo puede obtener una estimación insesgada de t_y ? - protesta el estadístico

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

De esta manera, ellos trabajan juntos en la elaboración de un plan de muestreo y diseñan un plan que asigna una probabilidad de inclusión de 99/100 a Sambo y probabilidades de 1/4900 al resto de la manada. Naturalmente, Sambo es seleccionado y el dueño del circo está feliz.

- ¿Cómo va a estimar t_y ? - pregunta el estadístico
- ¿Por qué? La estimación debería ser 50y, por supuesto - responde el dueño

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

-¡Oh! no, eso es incorrecto - responde el estadístico - recientemente, yo leí en un artículo de *Annals of Mathematical Statistics*, en donde se prueba que el estimador de Horvitz-Thompson es el único *estimador hiper admisible en la clase de todos los estimadores insesgados polinomiales generalizados*.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

- ¿Cuál sería la estimación de Horvitz-Thompson en este caso?
- pregunta el impresionado dueño - Dado que la probabilidad de inclusión de Sambo fue de $99/100$ - dice el estadístico - el estimador es $\hat{t}_y = 100/99y$ y no $50y$.
- ¿Y cuál sería nuestra estimación si el plan de muestreo hubiese seleccionado a Jumbo? (el elefante más pesado) - pregunta el incrédulo dueño.
- De acuerdo a lo que yo entiendo acerca del método de Horvitz-Thompson - dice el infeliz estadístico - el estimador de sería $\hat{t}_y = 4900y$ (donde y es el peso de Jumbo).

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

De esta forma, el estadístico perdió su empleo (y quizás se convirtió en profesor de estadística).

Basú, D. (1971). Foundations of Statistical Inference.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

*"Dada una muestra aleatoria seleccionada de una población...
Dadas n variables aleatorias independientes e idénticamente distribuidas"*

Kish (1996) afirma que cada palabra en el anterior enunciado es engañosa:

- Las muestras no son dadas, deben ser seleccionadas, asignadas o capturadas.
- El tamaño de la muestra no siempre es un número n fijo, en la mayoría de casos prácticos, es una variable aleatoria.
- Los datos no necesariamente siguen el supuesto de independencia ni de idéntica distribución.
- En muchas ocasiones no existe una sola población, sino que la muestra seleccionada es el resultado de una selección de subpoblaciones para las cuales se deben producir, no sólo una, sino muchas estimaciones.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Kish (1996) afirma que cada palabra en el anterior enunciado es engañosa:

- Las muestras no son dadas, deben ser seleccionadas, asignadas o capturadas.
- El tamaño de la muestra no siempre es un número n fijo, en la mayoría de casos prácticos, es una variable aleatoria.
- Los datos no necesariamente siguen el supuesto de independencia ni de idéntica distribución.
- En muchas ocasiones no existe una sola población, sino que la muestra seleccionada es el resultado de una selección de subpoblaciones para las cuales se deben producir, no sólo una, sino muchas estimaciones.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Kish (1996) afirma que cada palabra en el anterior enunciado es engañosa:

- Las muestras no son dadas, deben ser seleccionadas, asignadas o capturadas.
- El tamaño de la muestra no siempre es un número n fijo, en la mayoría de casos prácticos, es una variable aleatoria.
- Los datos no necesariamente siguen el supuesto de independencia ni de idéntica distribución.
- En muchas ocasiones no existe una sola población, sino que la muestra seleccionada es el resultado de una selección de subpoblaciones para las cuales se deben producir, no sólo una, sino muchas estimaciones.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Kish (1996) afirma que cada palabra en el anterior enunciado es engañosa:

- Las muestras no son dadas, deben ser seleccionadas, asignadas o capturadas.
- El tamaño de la muestra no siempre es un número n fijo, en la mayoría de casos prácticos, es una variable aleatoria.
- Los datos no necesariamente siguen el supuesto de independencia ni de idéntica distribución.
- En muchas ocasiones no existe una sola población, sino que la muestra seleccionada es el resultado de una selección de subpoblaciones para las cuales se deben producir, no sólo una, sino muchas estimaciones.

Suponga que la población de interés se puede escribir como $U = \{1, 2, \dots, N\}$ y que para estimar el total poblacional t_y , se selecciona una muestra probabilística s .

- Una estrategia de muestreo se define como la dupla $(p(s), \hat{t}_y)$; en donde $p(s)$ es un diseño muestral y \hat{t}_y es un estimador del parámetro de interés.
- Un estimador \hat{t}_y es lineal si puede ser escrito como

$$\hat{t}_y = \sum_s w_k y_k$$

- El ejemplo por excelencia de un estimador lineal es el estimador de Horvitz-Thompson, donde las ponderaciones $w_k = 1/\pi_k$ resultan ser iguales al inverso de la probabilidad de inclusión del elemento k .

Suponga que la población de interés se puede escribir como $U = \{1, 2, \dots, N\}$ y que para estimar el total poblacional t_y , se selecciona una muestra probabilística s .

- Una estrategia de muestreo se define como la dupla $(p(s), \hat{t}_y)$; en donde $p(s)$ es un diseño muestral y \hat{t}_y es un estimador del parámetro de interés.
- Un estimador \hat{t}_y es lineal si puede ser escrito como

$$\hat{t}_y = \sum_s w_k y_k$$

- El ejemplo por excelencia de un estimador lineal es el estimador de Horvitz-Thompson, donde las ponderaciones $w_k = 1/\pi_k$ resultan ser iguales al inverso de la probabilidad de inclusión del elemento k .

Suponga que la población de interés se puede escribir como $U = \{1, 2, \dots, N\}$ y que para estimar el total poblacional t_y , se selecciona una muestra probabilística s .

- Una estrategia de muestreo se define como la dupla $(p(s), \hat{t}_y)$; en donde $p(s)$ es un diseño muestral y \hat{t}_y es un estimador del parámetro de interés.
- Un estimador \hat{t}_y es lineal si puede ser escrito como

$$\hat{t}_y = \sum_s w_k y_k$$

- El ejemplo por excelencia de un estimador lineal es el estimador de Horvitz-Thompson, donde las ponderaciones $w_k = 1/\pi_k$ resultan ser iguales al inverso de la probabilidad de inclusión del elemento k .

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

En términos de estadísticas oficiales, el tipo de inferencia dominante después de la Segunda Guerra Mundial, fue la fundamentada en el diseño de muestreo (Neyman 1934, Cochran 1976) que descansa en la base de la aleatoriedad de las muestras y considera que las características de interés de cada individuo son parámetros poblacionales.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

A comienzos de los años setenta, Richard Royal, con la ayuda de otros coautores, quiso cambiar rotundamente esa tendencia con gran determinación. Él afirmó que la inferencia basada en el diseño, aunque no hace supuestos acerca de las probabilidades y parece ser no paramétrica y robusta, estaba sujeta a importantes defectos.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

La sugerencia de Royal fue radical: él propuso abandonar la inferencia basada en el diseño de muestreo, a favor de estimadores cuyas útiles propiedades (insesgamiento, consistencia, optimalidad, etc.) estuvieran definidas en términos del modelo predictivo apropiado.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Esto significa que conceptos como el sesgo y la varianza ya no están definidos como esperanzas a través de todas las posibles muestras, sino como promedios de las realizaciones de las unidades poblacionales (estén en la muestra o no) bajo el modelo predictivo establecido.

Desde el punto de vista de Royal, el proceso de aleatorización se convierte en irrelevante y propone que la muestra sea escogida a conveniencia (lo que en la práctica a veces significa escoger las unidades más grandes) puesto que las características de interés de cada individuo se consideran cantidades aleatorias.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

La inferencia basada en el diseño de muestreo varía radicalmente de la fundada en los modelos predictivos y quizás de cualquier otro modelo estadístico, porque está apoyada exclusivamente en las observaciones muestrales y no hace supuestos a priori; además, su dirección de análisis va en contravía con la dirección de la inferencia basada en modelos.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Kyburg (1987) afirma que: "la inferencia inversa procede de lo particular a lo general, la inferencia directa, de lo general a lo particular".

Desde este punto de vista, la inferencia basada en el diseño de muestreo es inversa puesto que al seleccionar una muestra, ésta sirve como base para obtener conclusiones acerca de la población; mientras que la fundamentada en modelos predictivos es directa porque supone que la población se comporta de acuerdo con la estructura probabilística de un modelo predictivo, por tanto, la muestra seleccionada debe comportarse igual.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú

La historia incompleta

Introducción

Definiciones básicas

Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

En términos de estadísticas oficiales, la tendencia moderna en algunos reconocidos institutos nacionales de estadística, es utilizar la inferencia basada en el diseño de muestreo (inferencia inversa) para obtener conclusiones sobre poblaciones o subgrupos poblacionales grandes; sin embargo, cuando se trata de realizar inferencia con subgrupos poblacionales pequeños (small areas) es conveniente emplear la inferencia basada en modelos poblacionales predictivos.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Se ha generado la creencia de que una muestra representativa es un modelo reducido de la población y de aquí se desprende un argumento de validez sobre la muestra: una buena muestra es aquella que se parece a la población, de tal forma que las categorías aparecen con las mismas proporciones que en la población” .

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

En ciertas ocasiones es fundamental "sobre representar" algunas categorías o incluso seleccionar unidades con probabilidades desiguales.

Tillé (2006) cita el siguiente ejemplo: suponga que el objetivo es estimar la producción de hierro en un país y que nosotros sabemos que el hierro es producido por dos compañías gigantes, con miles de empleados, y por cientos de pequeñas compañías, con pocos empleados.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

¿La mejor forma de seleccionar la muestra consiste en asignar la misma probabilidad a cada compañía? Claro que no; primero se averigua la producción de las grandes compañías; después, se selecciona una muestra de las compañías pequeñas con probabilidades desiguales y de algunas compañías grandes con probabilidad de inclusión forzosa; de forma que población y muestra difieren estructuralmente en el porcentaje de empresas pertenecientes a cada categoría.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

La muestra no debe ser un modelo reducido de la población; debe ser una herramienta usada para obtener estimaciones. Es así como el concepto de muestra representativa pierde peso, más aún, como una estrategia de muestreo es una dupla: diseño de muestreo (distribución de probabilidad sobre todas las posibles muestras) y estimador, el investigador debe preocuparse por ambos aspectos de esta dupla.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

La teoría de muestreo se ha ocupado de estudiar estrategias óptimas que permitan asegurar la calidad de las estimaciones; entonces, el concepto de representatividad debería estar asociado con las estrategias de muestreo y no sólo con las muestras.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Una estrategia se dice representativa si permite estimar un total poblacional exactamente, es decir, sin sesgo y con varianza nula. Si se utiliza el estimador de Horvitz-Thompson junto con un diseño de muestreo apropiado, esta estrategia es representativa sólo si, junto con la muestra seleccionada, el estimador reproduce algunos totales de la población.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Si para alguna muestra, se reproducen algunos totales de la población, se dice muestras balanceadas. Existen también, estimadores que brindan a la estrategia el calificativo de representativa, algunos de ellos son conocidos como de calibración.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Suponga que se utiliza el estimador de Horvitz-Thompson para estimar un parámetro de interés y que se planea un diseño muestral que induce probabilidades de inclusión desiguales; la regla de oro en este caso, exige que el vector de probabilidades de inclusión de primer orden sea proporcional a la característica de interés.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

El adjetivo proporcional debe entenderse como la similaridad en el comportamiento estructural del vector de probabilidades de inclusión y el vector de valores observados de la característica de interés.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Por ejemplo, a propósito de la repartición de la riqueza, se planean tres diseños de muestreo

- Asigna probabilidades de inclusión iguales a cada elemento de la población.
- Asigna mayores probabilidades de inclusión a los individuos que poseen menos y menores probabilidades de inclusión a los individuos que poseen gran cantidad de riqueza.
- Asigna probabilidades de inclusión mayores a los ricos y menores a los pobres (mejor diseño).

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Por ejemplo, a propósito de la repartición de la riqueza, se planean tres diseños de muestreo

- Asigna probabilidades de inclusión iguales a cada elemento de la población.
- Asigna mayores probabilidades de inclusión a los individuos que poseen menos y menores probabilidades de inclusión a los individuos que poseen gran cantidad de riqueza.
- Asigna probabilidades de inclusión mayores a los ricos y menores a los pobres (mejor diseño).

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Por ejemplo, a propósito de la repartición de la riqueza, se planean tres diseños de muestreo

- Asigna probabilidades de inclusión iguales a cada elemento de la población.
- Asigna mayores probabilidades de inclusión a los individuos que poseen menos y menores probabilidades de inclusión a los individuos que poseen gran cantidad de riqueza.
- Asigna probabilidades de inclusión mayores a los ricos y menores a los pobres (mejor diseño).

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Se acude al principio de representatividad sobre el cual se basa todo el andamiaje epistemológico de la inferencia en poblaciones finitas (Brewer 2002, pp 101-104). A pesar de la variación de todas la poblaciones, algunos individuos se representan a sí mismos y a algún otro conjunto de individuos.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Es por lo anterior que en estadística se utilizan ponderadores para representar a la población de interés y un ponderador natural es el inverso de la probabilidad de inclusión, por tanto, un individuo con una probabilidad de inclusión máxima igual a uno, sólo es capaz de representarse a sí mismo y a nadie más, puesto que el inverso de la unidad es la unidad.

Un individuo con una probabilidad de inclusión baja, se representará a sí mismo y a un conjunto grande de individuos.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

A veces, el estadístico puede utilizar procesos de estimación avanzados. Es muy importante que tenga en cuenta que la información auxiliar continua debe estar altamente correlacionada con la característica de interés, ya que estos procesos de estimación, en últimas, crean estimadores lineales cuyas ponderaciones no dependen directamente del diseño muestral.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Si la información auxiliar no está correlacionada con la característica de interés, las ponderaciones resultantes tampoco estarán correlacionadas con esta misma; por lo tanto, al final se tendrá una estrategia deficiente compuesta por un excelente diseño muestral y un pésimo estimador.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva
Encuestas electorales

A finales del siglo pasado, Smith (1976) hizo el siguiente llamado: **si los estadísticos teóricos hacen caso omiso al reto de enfrentar las encuestas multipropósito, entonces el vacío existente entre ellos y los estadísticos prácticos se hará cada vez más grande. El diseño y análisis de encuestas multivariantes debe ser una de las próximas áreas de mayor investigación...**

Motivación

Introducción

Diseño de muestreo

Estimación

Algunos escenarios prácticos

Comparación empírica

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Discusión

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva
Encuestas electorales

Dado que los institutos nacionales de estadística que manejan las cifras oficiales, con las cuales el gobierno decide acerca de sus políticas públicas, entre otros, realizan encuestas multipropósito, el estadístico debe preguntarse acerca del principio de representatividad en este ámbito.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva
Encuestas electorales

Si las encuestas que se planean responden al objetivo de la estimación de múltiples parámetros de interés, entonces el diseño de muestreo debería inducir probabilidades de inclusión que sean proporcionales al comportamiento estructural de todas las características de interés. Holmberg (2002) plantea una solución a este problema y encuentra un vector de probabilidades de inclusión óptimo.

El principio de representatividad

A. Gutiérrez

Motivación

- La fábula de Basú
- La historia incompleta

Introducción

- Definiciones básicas
- Un poco de historia

Diseño de muestreo

Estimación

- La regla de oro

Algunos escenarios prácticos

- Encuestas multipropósito

Inferencia predictiva

- Encuestas electorales

En la década de los setenta, varios autores dieron un viraje en su perspectiva epistemológica de la inferencia en poblaciones finitas, es así como Basú, Brewer, Godambe y Royal, entre otros, consideraron los modelos estadísticos (en sintonía con la estadística clásica Fisheriana) como los verdaderos fundamentos de la estimación e inferencia en poblaciones finitas.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Su trabajo se cimentó alrededor de la posibilidad de tener una inferencia que dependiera estrictamente del modelo propuesto y no tuviera nada que ver con el diseño de muestreo utilizado en la recolección de los datos. Como consecuencia, la atención se tornó alrededor de la estimación y se dejó de lado el muestreo por la relación existente o propuesta entre la característica de interés y las variables de información auxiliar.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

El camino que tomó la historia del muestreo fue, precisamente, la incorporación de las dos corrientes de pensamiento bajo una sola sombrilla, de modo que fue posible combinar la aleatorización clásica con una percepción más general de la relación de la característica de interés, con la información auxiliar disponible, y no hubo necesidad de sacrificar los principios basados en la aleatorización.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito

Inferencia predictiva

Encuestas electorales

Así nació la inferencia asistida por modelos, pero basada en la aleatorización (model assisted design-based inference), la cual es un matrimonio efectivo y tolerante que permite las ideas de la regresión junto con el paradigma de la aleatorización. Aunque los principios de la misma deben permanecer intocables porque las propiedades de los estimadores son evaluadas con respecto al mecanismo de probabilidad generado por la muestra y no con respecto a cualquier modelo asumido.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva

Encuestas electorales

En el último pulso electoral en Colombia, la tendencia de un empate técnico entre dos candidatos presidenciales era clara, continua y contundente. Apelando a la teoría estadística de la inferencia en poblaciones finitas basada en modelos poblacionales y reuniendo las observaciones de las distintas encuestas, se llega a conclusiones similares, pero esta vez respaldadas por las técnicas estadísticas sugeridas por Royal.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Se podría afirmar que existe evidencia estadística de que la votación de la primera vuelta de las elecciones presidenciales en Colombia no sería otra cosa que una reñida puja entre los dos candidatos; sin embargo, la realidad de las votaciones arrojó otro escenario muy distinto al que sugería la estadística basada en los modelos. La mayoría de las encuestas fue realizada en los principales municipios y ciudades del país de forma telefónica y unas pocas se hicieron por interceptación en la calle.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Sin embargo, a la mayoría de firmas encuestadoras se les olvidó que la subpoblación de colombianos con teléfono no representa a Colombia, que es un país pobre y con millones de personas en la indigencia.

Más aún, pareciese que obviaron que la carencia de teléfono no implica abstención en el voto. Luego, muchas personas de este gran conglomerado de ciudadanos votaron a favor del candidato oficialista y se presentó una victoria contundente.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva

Encuestas electorales

La situación anterior constituye una violación flagrante al principio de representatividad. Lo expuesto anteriormente, da pie para concluir, al menos empíricamente, que si el modelo sobre el cual se basa el proceso de estimación es errado, también lo serán las cifras resultantes de este proceso.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- El estadístico debe tener en cuenta el principio de representatividad en su práctica de planeamiento y ejecución de encuestas que arrojen estadísticas oficiales.
- En este ejercicio deben utilizarse estrategias de muestreo basadas en diseños de muestreo que induzcan probabilidades de inclusión que respeten el principio de la representatividad

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- El estadístico debe tener en cuenta el principio de representatividad en su práctica de planeamiento y ejecución de encuestas que arrojen estadísticas oficiales.
- En este ejercicio deben utilizarse estrategias de muestreo basadas en diseños de muestreo que induzcan probabilidades de inclusión que respeten el principio de la representatividad

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- El estadístico debe utilizar estimadores que hagan uso de estas probabilidades de inclusión proporcionales o que utilicen modelos que describan el comportamiento estructural de las características de interés; así, se asegura una estrategia de muestreo óptima.
- El diseño muestral y el estimador van en la misma vía del principio de representatividad

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- El estadístico debe utilizar estimadores que hagan uso de estas probabilidades de inclusión proporcionales o que utilicen modelos que describan el comportamiento estructural de las características de interés; así, se asegura una estrategia de muestreo óptima.
- El diseño muestral y el estimador van en la misma vía del principio de representatividad

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- Es un error utilizar un buen diseño muestral junto con estimadores que dejen de lado las ponderaciones inducidas por el diseño.
- Es un error no planear un excelente diseño muestral confiando en que un estimador avanzado arreglará estas falencias.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

- Es un error utilizar un buen diseño muestral junto con estimadores que dejen de lado las ponderaciones inducidas por el diseño.
- Es un error no planear un excelente diseño muestral confiando en que un estimador avanzado arreglará estas falencias.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

El estadístico debe involucrar el principio de la representatividad en la estrategia de muestreo, tanto en el diseño, como en el estimador.

El uso de la inferencia basada en modelos en este contexto debe ser limitado solamente a la estimación de áreas pequeñas, que en última instancia poseen un muy pequeño tamaño de muestra, que en algunos casos llega a ser nulo.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

La cualidad de representatividad en una estrategia no tiene sentido cuando no se cumple la regla de oro del muestreo; es decir cuando el comportamiento estructural de la característica de interés no es proporcional al comportamiento estructural de las probabilidades de inclusión π_k .

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Se simuló una población de tamaño $N = 1000$ de un modelo de superpoblación. Para este propósito, suponemos que la relación entre Y_k and \mathbf{x}_k puede ser descrita mediante un modelo ξ , tal que $Y_k = \mathbf{x}_k\beta + \varepsilon_k$ y para todo $k = 1, \dots, N$.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Se supone que $\varepsilon_1, \dots, \varepsilon_N$ constituyen un conjunto de variables aleatorias con distribución normal tal que:

$$\begin{aligned} E_{\xi}(\varepsilon_k) &= 0 \\ \text{Var}_{\xi}(\varepsilon_k) &= \sigma_k^2 = \sigma^2 \times x_k^2. \end{aligned} \tag{1}$$

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

El valor del parámetro β se fija igual a treinta. Además, se asumió que los ε_k son independientes, con estructura heteroscedástica. En cada corrida de la simulación se tomaron muestras aleatorias simples sin balancear de tamaño $n = 100$, $n = 300$ y $n = 500$ y muestras balanceadas de igual tamaño.

Así, se definieron las siguientes estrategias de muestreo para la estimación del total poblacional, t_y :

- $(MAS, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo balanceado.
- $(MAS, \hat{t}_{y,cal})$: el estimador de calibración bajo muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,cal})$: el estimador de calibración con un diseño de muestreo balanceado.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Así, se definieron las siguientes estrategias de muestreo para la estimación del total poblacional, t_y :

- $(MAS, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo balanceado.
- $(MAS, \hat{t}_{y,cal})$: el estimador de calibración bajo muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,cal})$: el estimador de calibración con un diseño de muestreo balanceado.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Así, se definieron las siguientes estrategias de muestreo para la estimación del total poblacional, t_y :

- $(MAS, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo balanceado.
- $(MAS, \hat{t}_{y,cal})$: el estimador de calibración bajo muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,cal})$: el estimador de calibración con un diseño de muestreo balanceado.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Así, se definieron las siguientes estrategias de muestreo para la estimación del total poblacional, t_y :

- $(MAS, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,\pi})$: el estimador clásico con un diseño de muestreo balanceado.
- $(MAS, \hat{t}_{y,cal})$: el estimador de calibración bajo muestreo aleatorio simple.
- $(BAL, \hat{t}_{y,cal})$: el estimador de calibración con un diseño de muestreo balanceado.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

El proceso se repitió $B = 1000$ veces y en cada simulación, el desempeño de un estimador \hat{t}_y fue evaluado usando su sesgo relativo, y su eficiencia relativa, definidas por, respectivamente, como:

$$SR(\hat{t}_y) = B^{-1} \sum_{b=1}^B \frac{\hat{t}_{y,b} - t_y}{t_y} \quad (2)$$

$$ER(\hat{t}_y) = \frac{ECM(\hat{t}_y)}{ECM_{MAS}(\hat{t}_y, \pi)}, \quad (3)$$

donde

$$ECM(\hat{t}_y) = B^{-1} \sum_{b=1}^B (\hat{t}_{y,b} - t_y)^2 \quad (4)$$

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Teniendo en cuenta que $\hat{t}_{y,b}$ se calculó en la b -ésima muestra simulada, se nota que grandes valores para $ER (> 1)$ representan una baja eficiencia de la estrategia $(p(\mathbf{s}), \hat{t}_y)$ en comparación a la estrategia $(MAS, \hat{t}_{y,\pi})$. Al contrario, valores cercanos menores a uno o cercanos a cero para $ER (< 1)$ representan una alta eficiencia de la estrategia.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Por otro lado, se consideraron varios escenarios de correlación entre la característica de información auxiliar y la característica de interés. Es así como, cambiando los valores de σ^2 , se crearon tres escenarios: el primero, en donde la correlación es casi perfecta e igual a 0.99, el segundo en donde la correlación es de 0.8, y el último en donde la correlación es de 0.2.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

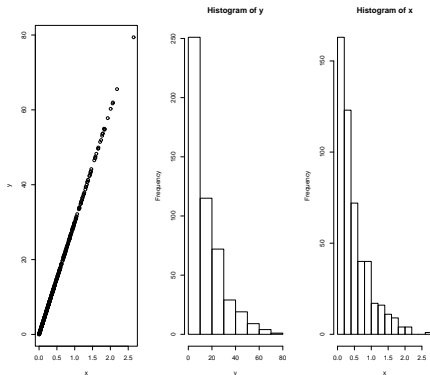


Figura: Diagrama de puntos de y contra x , histograma de y e histograma de x para $\text{corr}(x, y) = 0,99$.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

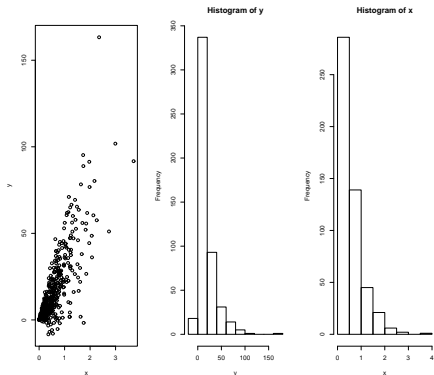


Figura: Diagrama de puntos de y contra x , histograma de y e histograma de x para $\text{corr}(x, y) = 0,80$.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

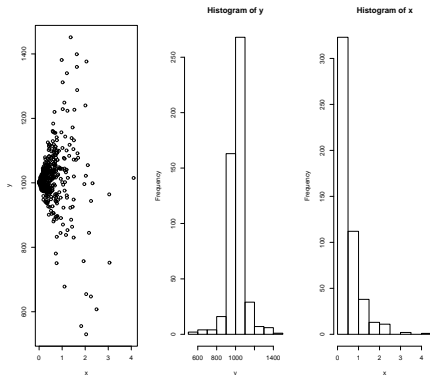


Figura: Diagrama de puntos de y contra x , histograma de y e histograma de x para $\text{corr}(x, y) = 0,20$.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Con respecto al sesgo relativo, todos los resultados estuvieron muy cercanos a cero. Lo anterior es natural pues el estimador de Horvitz-Thompson es insesgado bajo cualquier diseño sin reemplazo y el estimador de calibración es asintóticamente insesgado bajo cualquier diseño de muestreo sin reemplazo.

Cuadro: *Eficiencia relativa de las estrategias de muestreo para $\text{corr}(x, y) = 0,99$.*

n	Muestreo aleatorio simple		Muestreo balanceado	
	$\hat{t}_{y,\pi}$	\hat{t}_{cal}	$\hat{t}_{y,\pi}$	$\hat{t}_{y,cal}$
10	1.000	0.335	0.495	0,345
50	1.000	0.344	0.365	0.336
200	1.000	0.388	0.401	0.380

Cuadro: *Eficiencia relativa de las estrategias de muestreo para $\text{corr}(x, y) = 0,80$.*

n	Muestreo aleatorio simple		Muestreo balanceado	
	$\hat{t}_{y,\pi}$	\hat{t}_{cal}	$\hat{t}_{y,\pi}$	$\hat{t}_{y,cal}$
10	1.000	0.871	0.965	0.908
50	1.000	0.923	0.892	0.889
200	1.000	0.835	0.802	0.799

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
 La historia incompleta

Introducción

Definiciones básicas
 Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
 Inferencia predictiva
 Encuestas electorales

Cuadro: *Eficiencia relativa de las estrategias de muestreo para $\text{corr}(x, y) = 0,20$.*

n	Muestreo aleatorio simple		Muestreo balanceado	
	$\hat{t}_{y,\pi}$	\hat{t}_{cal}	$\hat{t}_{y,\pi}$	$\hat{t}_{y,cal}$
10	1.000	1.075	1.115	1.093
50	1.000	0.998	1.027	1.020
200	1.000	1.008	1.063	1.065

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

En los escenarios con alta correlación, se nota que bajo muestreo aleatorio simple, el estimador de calibración es siempre más eficiente que el estimador de Horvitz-Thompson. Asimismo, bajo muestreo balanceado, el estimador de calibración es siempre más eficiente que el estimador de Horvitz-Thompson.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Cuando la correlación es casi perfecta, la ganancia en eficiencia de las estrategias $(BAL, \hat{t}_{y,\pi})$, $(MAS, \hat{t}_{y,cal})$ y $(BAL, \hat{t}_{y,cal})$ es muy alta, con respecto a $(MAS, \hat{t}_{y,\pi})$. Sin embargo, cuando la estructura de correlación baja, la ganancia en eficiencia no es tan grande.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

En el escenario en donde no se tiene una buena correlación, el comportamiento es diferente. Bajo muestreo aleatorio simple, al utilizar un estimador de calibración se aumenta la variación y por tanto se concluye que la distribución del estimador es poco precisa.

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Cuando se tiene este escenario, la mejor estrategia es $(MAS, \hat{t}_{y,\pi})$. Lo anterior es concluyente debido a que en la vida real, los dos primeros escenarios son muy poco frecuentes y las características de información auxiliar raras veces presentan un comportamiento estructural proporcional al de la característica de interés

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Cuando se utiliza un diseño de muestreo sin reemplazo proporcional al tamaño de una característica de información auxiliar (πPT), el estimador de calibración coincide plenamente con el estimador de Horvitz-Thompson. Las probabilidades de inclusión de primer están dadas por

$$\pi_k = n \frac{x_k}{t_x} \quad (5)$$

Por lo tanto, el estimador de Horvitz-Thompson es tal que

$$\hat{t}_{x,\pi} = \sum_{k \in S} \frac{x_k}{\pi_k} = t_x$$

y como consecuencia, el estimador de calibración toma la siguiente forma

$$\hat{t}_{y,cal} = \hat{t}_{y\pi} + (t_x - t_{x\pi})\hat{B} = \hat{t}_{y\pi}.$$

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

Cuadro: *Eficiencia relativa de estrategias de muestreo para $\text{corr}(x, y) = 0,20$ con probabilidades de inclusión desiguales y proporcionales al tamaño de x .*

n	$N\bar{y}_S$	$\hat{t}_{y,\pi}$	\hat{t}_{cal}
10	1.000	1521.9	587.3
50	1.000	1125.1	515.3
200	1.000	845.6	651.1

El principio de representatividad

A. Gutiérrez

Motivación

La fábula de Basú
La historia incompleta

Introducción

Definiciones básicas
Un poco de historia

Diseño de muestreo

Estimación

La regla de oro

Algunos escenarios prácticos

Encuestas multipropósito
Inferencia predictiva
Encuestas electorales

El cociente de las varianzas de estas estrategias con las varianza de una estrategia con diseño simple y con el estimador de expansión $N\bar{y}_S = N/n \sum_S y_k$, denotada como $(MAS, N\bar{y}_S)$, es muy grande, como se puede ver en la tabla 4. Por lo tanto, en estos casos es mejor abstenerse de utilizar estrategias sofisticadas y utilizar la estrategia más simple.

Agradecimientos

Muchas gracias !!!

<http://www.gutierrezandres.com/blog>